A comprehensive evaluation of rhythm monitoring strategies for the detection of atrial fibrillation recurrence: insights from 647 continuously monitored patients and implications for monitoring after therapeutic interventions

Article  (Unspecified)
A Comprehensive Evaluation of Rhythm Monitoring Strategies for the Detection of Atrial Fibrillation Recurrence: Insights From 647 Continuously Monitored Patients and Implications for Monitoring After Therapeutic Interventions

Efstratios I. Charitos, MD; Ulrich Stierle, MD; Paul D. Ziegler, MS; Malte Baldewig, MD; Derek R. Robinson, MA, MSc, DPhil, CStat; Hans-Hinrich Sievers, MD; Thorsten Hanke, MD

Background—Intermittent rhythm monitoring (IRM) to detect atrial fibrillation (AF) recurrence is employed to evaluate the success of therapeutic interventions. In a large population of patients with continuous monitoring (CM), we investigated the sensitivity of various frequencies and durations of IRM strategies on the detection of AF recurrence, the dynamics behind AF recurrence detection, and we describe measures to evaluate temporal AF recurrence.

Methods and Results—Rhythm histories of 647 patients (mean AF burden, 0.12 ± 0.22; median, 0.014; 687 patient-years) with implantable CM devices were reconstructed and analyzed. With the use of computationally intensive simulation, the sensitivity of IRM of various frequencies and durations on the identification of AF recurrence was evaluated. Prolonged-duration IRM was superior to shorter IRM (P < 0.0001). However, even with aggressive IRM strategies, AF recurrence was not detected in a great proportion of patients. The temporal AF burden aggregation (AF density) was directly related to IRM sensitivity (P < 0.0001). Even at similar AF burdens, patients with high-density AF required higher-frequency or prolonged-duration IRM to achieve the same sensitivity as in low-density AF (P < 0.0001). Patients with high-density, low-burden AF benefit the most from CM for detection of AF recurrence.

Conclusions—IRM follow-up is significantly inferior to CM. IRM strategies will not identify AF recurrence in a great proportion of patients at risk. Temporal AF characteristics play a significant role in AF recurrence detection with the use of IRM. For the scientific, evidence-based evaluation of AF treatments, CM should be strongly recommended. Prospective studies are required to evaluate whether CM to guide clinical management can also improve patient outcomes.

Clinical Trial Registration—URL: http://www.clinicaltrials.gov. Unique identifier: NCT00806689.

Key Words: atrial fibrillation ■ atrial fibrillation arrhythmia ■ rhythm monitoring ■ rhythm recorders

Detection of atrial fibrillation (AF) recurrence after therapeutic interventions has until now been based primarily on the results of intermittent (“snapshot”) rhythm monitoring (IRM). This approach has significant limitations in terms of sensitivity because the monitoring period of these examinations is limited; however, the results of these examinations are used to draw inferences on the success of ablation procedures or pharmacological strategies. It has been shown that reliance on symptoms or office ECGs overestimates the success rate of the ablation procedures and will misclassify patients who have recurrence of paroxysmal AF as being in sinus rhythm.1-4 The current consensus on AF monitoring recommends at least two 24-hour Holter monitor (HM) examinations annually for the detection of AF recurrence after ablation procedures.5-7 which, however, has also been shown to underdetect AF recurrence and thus overestimate procedural success.1,8 Reliable and accurate detection of AF recurrence is thus of special importance for the evaluation of pharmacological or ablation therapies as well as when decisions on changes in anticoagulation or antiarrhythmic therapy are to be made.

Editorial see p 791

Clinical Perspective on p 814

With the introduction of implantable, leadless rhythm recorders, continuous monitoring (CM) has been proposed...
recently for more accurate detection of AF recurrence after ablation procedures or novel pharmacological strategies. As expected, the sensitivity of these diagnostic modalities is close to 100% and, as such, is much superior to that of the currently employed 24-hour HM strategies. However, the implantation of a CM device in all patients may seem unrealistic. This led to the belief that intensifying noninvasive IRM either with greater frequency or with longer duration (7- or 30-day monitoring) or both may lead to better and more reliable detection of AF recurrence. 

Recently, data from CM patients have been used to correlate the recurrence of AF with the risk for thromboembolic complications. The advance of this technology has not only provided better insight into the rhythm evaluation of these patients but has also initiated a change of mentality regarding AF from a qualitative (“yes/no”) to a quantitative (amount of AF) approach. The AF burden, defined as the proportion of the total monitored time a patient is in AF, has been utilized in evaluating the risk for thromboembolism in patients with AF. However, even these measures present a rather static approach in the evaluation of the dynamic incidence of AF recurrence, partially ignoring the temporal dispersion or temporal aggregation of the AF episodes and AF burden.

The aim of the present study was 2-fold: (1) to identify the sensitivity of IRM strategies of various frequencies and durations on the accurate detection of AF recurrence with the use of data collected from patients with heart rhythm CM devices and (2) to investigate the dynamics behind AF recurrence detection and propose measures to evaluate the temporal pattern of AF recurrence.

### Methods

Data acquired from 647 patients monitored with a CM device (Reveal XT 9529, n=73; AT500 pacemaker, n=574; Medtronic, Inc, Minneapolis, MN) were analyzed. Demographics and patient characteristics are summarized in Table 1. All patients provided informed consent for the data collection and use, and the study was approved by the local ethics committee.

With the use of data from the CM device, all AF episodes were examined. Isolated AF episodes of <5 minutes duration that could present artifacts were disregarded, and patients with no documented AF episode were regarded as AF free (n=174). The complete rhythm history of every patient was reconstructed (Figure 1). Thereafter, we calculated the probability of successful identification of AF recurrence by IRM of various durations (eg, 24 hours and 7, 14, and 30 days) and frequencies for every patient. Computationally intensive simulation was employed to simulate in every patient all possible IRM strategies of various durations and frequencies to draw inferences on the sensitivity of the monitoring strategy and burden characteristics of the patients being monitored. Sensitivity was defined as the proportion of patients correctly identified as having AF recurrence with the simulated IRM strategy to the true number of patients having AF recurrence (identified from CM). AF burden was defined as the proportion of the time in AF in relation to the total monitored time. The simulation trials were performed as follows: After reconstructing the rhythm history of every patient j, monitored for a total of g days, we defined the sample space \( \Omega_g = \{1, 2, \ldots, g-k+1\} \) to be the set of possible days that a k-day intermittent monitoring could be started (for \( k=1, 2, 7, 14, \) and 30). A k-day monitoring starting on day \( t \in \Omega_g \) therefore included the following associated monitored days: \( t, t+1, \ldots, t+k-1 \). To simulate \( n \) independent k-day monitorings of patient j, \( n \) elements were selected at random from \( \Omega_g \), except that elements were rejected if their monitored days intersected with the monitored days of previously selected elements. AF was deemed to have been successfully identified if it was observed in at least 1 of the \( n \) sets of monitored days. This was performed for all patients of the study population, for monitoring durations of \( k=1, 2, 7, 14, \) and 30 days, and for strategies...
Figure 1. Reconstruction of the rhythm history of patients A and B. Both patients have similar atrial fibrillation (AF) burden (AF burden for patient A = 0.22; AF burden for patient B = 0.21). The vastly different temporal aggregation of the AF episodes and AF burden in these patients is responsible for the different sensitivities of intermittent AF monitoring modalities in these patients. For patient A, the probability of identification of AF recurrence in a 24-hour and 30-day monitor is \( Pr(24h|A) = 0.23 \) and \( Pr(30d|A) = 0.26 \). In this case, a random 30-day monitoring increases the sensitivity by only 3% from that which random 24-hour Holter monitoring (HM) would provide. For patient B, the respective probabilities are \( Pr(24h|B) = 1 \) and \( Pr(30d|B) = 1 \). The numbers 1 and 2 denote the time points of 2 random 24-hour HM tests, whereas the numbers 3 and 4 denote the time points of 2 random 30-day HM tests.

Sensitivity of IRM

Quantitative AF characteristics are displayed in Figure 3. Figure 4 displays the results for the sensitivity of random 24-hour and 7-, 14-, and 30-day monitorings at various frequencies to detect AF recurrence for the entire patient population. Prolonged IRM was significantly superior to shorter IRM (\( P < 0.0001 \) for all comparisons). However, a four 24-hour HM strategy had a sensitivity of only 52%, thus failing to identify AF recurrence in almost half of the patients with proven AF recurrence. Even with a theoretical strategy of three 30-day HM tests (for a total monitoring duration of 90 days per year), the sensitivity did not exceed 82%, and thus a nonnegligible proportion of patients with AF would be misclassified (Figure 4). However, the increased sensitivity offered by prolonged IRM durations comes at a cost of disproportionate increase in the required monitored time to achieve that level of sensitivity. For a sensitivity of 0.65, 30 days of monitoring would have been required with a 30-day
HM strategy (one 30-day HM test), whereas for the same sensitivity, 7 monitored days would have been required with a 24-hour HM strategy (seven 24-hour HM tests) (Figure 4, solid horizontal black line).

Inferences Drawn From a Series of Negative IRM
Using computationally intensive simulation, we sought to evaluate the inferences that can be drawn from a series of negative IRM on the patients’ AF burden. Although IRM strategies cannot determine the precise AF burden, reliable inferences can be obtained (for example, a patient with 4 negative random 30-day HM tests is highly unlikely to have a burden of >80%). These probabilities were derived from our patient population, taking into consideration not only the amount of AF in each patient but also the temporal characteristics of AF, and are displayed in Figure 5. A series of 4 negative random 30-day HM tests suggests with 90% confidence that this patient’s AF burden is 17%. Similarly, 8 negative 24-hour HM tests would be required to achieve the same level of confidence regarding the patient’s AF burden. IRM of shorter duration requires a higher number of examinations to achieve the same level of confidence as IRM of longer durations (Figure 5).

Influence of AF Burden and AF Density on Effectiveness on IRM
Using the reconstructed rhythm histories of our patients, we attempted to evaluate the quantitative as well as temporal AF characteristics that drive the sensitivity differences between IRM of various durations and frequencies. Burden density was directly related to the sensitivity of the different monitoring durations (Figure 6). Even with similar AF burdens (burden of low-density group: mean, 0.12±0.11; median, 0.07; quartile 1, 0.01; quartile 3, 0.18; burden of high-density group: mean, 0.13±0.14; median, 0.06; quartile 1, 0.01; quartile 3, 0.19; \( P=0.36 \)), patients with low-density AF (AF density <0.5) achieve higher sensitivities for AF recurrence detection compared with patients with high-density AF (AF density >0.5; \( P<0.0001 \) for all comparisons of all monitoring duration sensitivities [24-hour, 7-day, 14-day, 30-day HM] between high- and low-density groups). In patients with high-density AF, AF recurrence is much more difficult to
capture with the use of IRM than in patients with low-density AF, even at similar AF burdens.

The superiority of CM in comparison to IRM was found to be significantly related to the AF burden and the AF density (Table 2). When CM is compared with 24-hour HM, the lower the AF burden and the higher the AF burden density, the greater is the gain from CM for AF recurrence detection (Figure 7, left). Patients with low-burden, high-density AF profit the most from CM. When CM was compared with a 30-day HM strategy, the gain obtained with CM was also found to be significantly related to the AF density and the AF burden. The higher the AF density and the lower the AF burden, the higher is the detection probability gain from the CM strategy (Figure 7, right).

**Discussion**

This study provides insights into the sensitivity of various monitoring strategies for the detection of AF recurrence with the use of data from a large patient population monitored with CM devices. Although 2 different monitoring devices were employed, AF detection in both devices was similar. The aim of our study was not to evaluate the AF detection process of the devices, which has been extensively validated previously, but to evaluate the characteristics and dynamics of AF recurrence detection after AF episode registration.

**Judgment of Therapeutic Success Based on IRM**

Our group and other groups have previously evaluated sensitivity differences between IRM and CM by comparing the information obtained from CM devices with actual IRM performed in the same patients. However, these methods have limitations because the time point of the IRM plays a major role in the evaluation of these methods. As depicted in Figure 1, in patient A, a 24-hour HM test performed at time point 1 will be negative and patient A will be regarded as “AF free,” whereas the same examination performed some days earlier at time point 2 would have identified AF recurrence. The results of such studies comparing continuous versus actual IRM are severely influenced by the choice of time points when the IRM is performed. Because in most of these studies only a limited number of IRM tests were compared...
against CM, chance has an immeasurable effect on the results of such studies, and any inferences drawn from these results will be problematic. In the present study, the reconstruction of the rhythm history of every patient allows us to evaluate the true probability of AF recurrence detection of any random IRM of any duration in every patient, and computationally intensive simulation can draw inferences on the sensitivities and success rates of IRM of any duration and frequency.

Figure 4 presents the results of this approach in our patient collective, depicting the sensitivities of the 4 most common IRM durations (24-hour and 7-, 14-, and 30-day HM) at various monitoring frequencies. In our patient population, a monitoring strategy with four 24-hour HM tests would have a sensitivity of only 52%, thus failing to identify recurrence of AF in almost half of these patients. This low sensitivity becomes even more striking if one considers that the flipping of a fair coin would have provided the same sensitivity for the detection of AF recurrence in these patients. The direct implication of these results is that the evaluation of pharmacological or interventional therapies when a 24-hour HM is employed as a monitoring strategy should be interpreted cautiously because a great proportion of patients will be misclassified, AF recurrence will be underdiagnosed, and therapeutic success will be overestimated. Similarly, a single, random 30-day HM in our patient population will have a sensitivity of 63%, and thus the evaluation of the success of an ablation procedure with the use of a single, postinterventional, 30-day HM would have a sensitivity of only 63% (slightly better than the sensitivity the flipping of a fair coin would provide). Even in the theoretical scenario of using three 30-day HM tests (for a total of 90 monitored days per year) would have a sensitivity of 82%; however, such a strategy seems unrealistic because it has been shown that complex or prolonged IRM strategies severely affect patient compliance. To obtain the sensitivities of the various IRM strategies presented in this report, a 100% patient compliance is assumed; however, this is rarely the case in everyday clinical practice because even short-duration IRMs pose a significant compromise in the patients’ quality of life, and this has been shown to lead to monitoring discontinuation.

Table 2. Linear Regression Analysis of Effect of Burden and Burden Density on the Probability Gain of AF Recurrence Detection of 30-Day and 24-Hour HM vs Continuous Monitoring

<table>
<thead>
<tr>
<th>Factor</th>
<th>Coefficient</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gain in AF recurrence detection: continuous vs 24-h HM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF burden</td>
<td>-5.76</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>AF burden density</td>
<td>2.61</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Gain in AF recurrence detection: continuous vs 30-d HM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF burden</td>
<td>-4.08</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>AF burden density</td>
<td>4.84</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

AF indicates atrial fibrillation; HM, Holter monitoring. To restore normality, the log of the probability gain was used as dependent variable.

Figure 6. Sensitivity of various intermittent rhythm monitoring durations at various frequencies in patients with high (>0.5) and low (<0.5) atrial fibrillation (AF) burden density. Both high- and low-density AF groups had similar AF burdens (burden of low-density group: mean, 0.12±0.11; median, 0.07; quartile 1, 0.01; quartile 3, 0.18; burden of high-density group: mean, 0.13±0.14; median, 0.06; quartile 1, 0.01; quartile 3, 0.19; P=0.36). Low-density AF patients achieve higher AF recurrence detection sensitivity with less frequent monitoring than patients with high-density AF. In patients with high-density AF (blocks of AF), a significantly (P<0.0001) higher monitoring frequency is required to detect recurrence of AF. Only patients with documented AF recurrence were included. HM indicates Holter monitoring.
intermittent monitoring altogether, a great part or even the majority of patients will be misclassified, and therefore the scientific evaluation of therapeutic interventions for AF becomes problematic.

Semiquantitative Estimation of AF Burden Based on Negative IRM

Although IRM may fail to detect AF recurrence in a great proportion of patients and cannot provide the level of information a CM strategy does, results of serial IRM may guide patient management in a semiquantitative way when CM is not possible because of either limited availability, patient compliance, or cost considerations. Such an evaluation is depicted in Figure 5. A series of negative IRM, although may underestimate qualitative AF detection per se, can however guide estimation of the potential AF burden level in patients prone to AF recurrence such as patients after ablation procedures, or patients on rhythm control strategies. In our large population of CM patients, 4 random negative 24-hour HM tests indicate, with a confidence of 90%, an AF burden of up to 23%. A higher number of serial negative IRM tests or negative IRM tests of greater duration can pinpoint the potential AF burden to lower levels. However, it is important to note that even average burdens that low have been shown to be a risk factor for thromboembolic complications,13–16,21

AF Characteristics Influencing the Success of IRM

A major variable driving the sensitivity in detecting AF recurrence is not only the AF burden but the dispersion of it over time. AF density, as a measure of this dispersion, proved to be a major factor that influences the sensitivity of IRM at various frequencies and durations. Although patients A and B of Figure 1 have very similar amounts of AF burden over the same observation time, the probability of AF recurrence detection is vastly different because of the different temporal AF burden dispersion. In patient B (Figure 1), because of the even dispersion of AF burden, prolonged-duration IRM (30-day HM) is not superior to shorter-duration IRM (24-hour HM), the probability of AF recurrence detection of both IRM tests is 1, and the sensitivity of both strategies is 100%. In patient A (Figure 1), however, because of the high AF density, a random 30-day HM test offers only a negligible 3% increase in probability of AF recurrence detection compared with a random 24-hour HM test. In the whole patient cohort (Figure 6), the AF density appears to have a major and statistically significant influence on the sensitivity of AF detection with the use of IRM strategies. In patients with high-density AF, the probabilities of AF recurrence detection and the sensitivities of IRM strategies are significantly inferior to those of patients with same level of AF burden but low AF density (Figure 6).

Implications for Clinical Outcomes, Patient Management, and Current Knowledge

Although the level of AF burden has been shown to be a significant risk factor for thromboembolic events,13–16 to the best of our knowledge, no study has taken into consideration the temporal characteristics of AF burden development. It is conceivable that the temporal distribution of AF burden may also play a role in the general thromboembolic risk that patients with AF face. Patients A and B (Figure 1), although spending the same time in AF, may have a different thromboembolic risk because of the different temporal distribution of their AF burden. Larger studies are required to evaluate this hypothesis, and prospective studies are required to evaluate the clinical impact of the combined quantitative and temporal AF characteristics on patient outcome.

The data presented may have some clinical implications. The results of our study depict the major limitations of IRM not only for management of patients but also for evaluation of the results of pharmacological as well as interventional therapies for AF. From a clinical perspective, accurate patient management can only be obtained with confidence with CM. From a scientific perspective, our evaluation until now of the success of pharmacological and interventional procedures based on IRM seems problematic in light of the results presented in this study. The accurate evaluation of therapies for AF in the era of evidence-based medicine mandates the use of CM.

It is conceivable that the widespread use of CM or the results from large studies of CM patients may require us to reevaluate our definitions of success or failure of therapeutic intervention for AF. It may very well be that the current definition of success of a therapeutic intervention only in terms of complete absence of AF recurrence might be inadequate and that a quantitative approach to AF recurrence may in the future require us to critically revisit our current knowledge and reevaluate our therapeutic interventions and
strategies, our results, and our patient management decision-making process. Whether the use of CM to guide clinical management can also improve patient outcomes remains uncertain at this point and must be investigated in large randomized controlled prospective studies.

Limitations
This study was underpowered to investigate the clinical impact of temporal AF characteristics on patient outcomes. We believe that this should be the scope of large prospective studies or registries of CM patients over sufficient follow-up time. AF episodes of patients primarily implanted with a CM device because of atrial tachycardia or atrial flutter have been included in the analysis. In our experience, although patients are typically classified in 1 distinct rhythm type (atrial flutter, AF, or atrial tachycardia), CM reveals that these patients actually tend to experience multiple types of atrial rhythms. Although 2 different devices were employed in the population of this study, previous studies have extensively investigated the accuracy of AF detection of these devices (sensitivity, ≥98%; specificity, ≥91%). Additionally, we tried to manually inspect all available episodes and, in cases of uncertainty or dispute, we discarded isolated, short (<5 minutes) episodes that could represent artifacts.

Conclusion
Patient follow-up with either short- or long-duration IRM is significantly inferior to CM for the detection of AF recurrence. IRM monitoring strategies will not identify AF recurrence in a great proportion of patients at risk. With short-duration, low-frequency IRM strategies, chance has an immeasurable effect on the evaluation of AF recurrence. Thus, results of pharmacological or invasive interventions evaluated with IRM should be considered with caution. Temporal AF characteristics play a major role in AF recurrence detection with the use of IRM strategies and may influence clinical parameters and outcomes in AF patients. For the scientific, evidence-based evaluation of AF treatments, CM should be strongly recommended.

Acknowledgments
We are thankful to Katrin Meyer and Martina Schröder for secretarial support and to Jana Paise, Anja Paap, and Tobias Frin for data documentation support.

Disclosures
Drs Charitos, Sterle, Baldewig, Robinson, and Sievers have no conflict of interest to disclose. Paul D. Ziegler is a full-time employee of Medtronic (>$10,000 USD) and stockholder of Medtronic (>10,000 USD). Dr Hanke discloses modest lecture honoraria (<10,000 USD) from Medtronic.

References
Patient follow-up with either short- or long-duration intermittent rhythm monitoring is significantly inferior to that with continuous rhythm monitoring for the detection of atrial fibrillation (AF) recurrence. Intermittent rhythm monitoring strategies fail to identify AF recurrence in a great proportion of patients at risk. Even with feasible, aggressive intermittent monitoring strategies (such as quarterly 24-hour Holter monitoring), AF recurrence will not be identified in a significant proportion of patients. Because of the low sensitivity of intermittent rhythm monitoring, the results of pharmacological or invasive interventions for AF when evaluated with intermittent monitoring should be considered with caution because chance has an immeasurable effect on the detection of AF recurrence, and thus a great proportion of patients will be misclassified, AF recurrence will be underdiagnosed, and therapeutic success will be overestimated. Novel quantitative and temporal AF characteristics (AF burden, AF density) play a major role in AF recurrence detection with the use of intermittent monitoring strategies and may influence clinical parameters and outcomes in AF patients. For accurate patient management, as well as for the scientific, evidence-based evaluation of AF treatments, continuous rhythm monitoring should be strongly recommended.

Go to http://cme.ahajournals.org to take the CME quiz for this article.